Redescribing Intonational Categories with Functional Data Analysis

Margaret Zellers[1], Michele Gubian[2], Brechtje Post[1]

[1]Research Centre for English & Applied Linguistics, University of Cambridge, UK
[2]Centre for Language & Speech Technology, Radboud University, Nijmegen, the Netherlands
mkz21@cam.ac.uk, M.Gubian@let.ru.nl, bmbp@cam.ac.uk

Abstract

Intonational research is often dependent upon hand-labeling by trained listeners, which can be prone to bias or error. We apply tools from Functional Data Analysis (FDA) to a set of fundamental frequency (F0) data to demonstrate how these tools can provide a less theory-dependent way of investigating F0 contours by allowing statistical analyses of whole contours rather than depending on theoretically-determined "important" parts of the signal. The results of this analysis support the predictions of current intonational phonology while also providing additional information about phonetic variability in the F0 contours that these theories do not currently model.
Index terms: fundamental frequency, intonation, Functional Data Analysis

1. Introduction

In the last few decades, much of intonational phonology has made use of an autosegmental-metrical analysis of intonation. In English, the primary descriptive intonational theory is the ToBI system (Silverman et al. [1]), based upon the intonational grammar developed by Pierrehumbert and colleagues (Pierrehumbert [2]; Pierrehumbert & Beckman [3]). This system assumes a grammar of legal combinations of high (H) and low (L) tones, which combine with each other to create the legal intonational tunes of English. An alternative description in a similar framework, focused specifically on British English intonation, has been provided by Gussenhoven ([4],[5]). One assumption of these types of models is that high and low tones are the main elements which contrast, as opposed to previous intonational theories (cf. Cruttenden [6], 't Hart et al. [7]) particularly in the British tradition, which have assumed that pitch movements, rather than tones/targets, are the minimal elements of intonation. By framing an analysis of intonation in terms of a binary contrast between high and low, the autosegmental tradition and specifically the analyses above aim to make a phonetically transparent phonological analysis of intonation, taking into account the meaningful aspects of the speech signal while ignoring extraneous variation.

One difficulty faced by intonational research is that analysis based on the work of trained listeners doing hand annotation is always prone to bias. Annotators must make decisions about which parts of the fundamental frequency (F0) contour are important and which are not: for example, deciding whether a pitch contour contains a peak or a plateau, which has been shown to be relevant in terms of the interpretation of peak timing (Knight [8]). This is especially problematic because inter-annotator agreement can vary dramatically (cf. Syrdal & McGory [9], who found that pairwise agreement on labeling met or exceeded 50% for less than half of the ToBI pitch accents).

Recently, some researchers (Gubian et al. [10]) have been investigating the use of Functional Data Analysis (FDA) as a way of getting around some of the difficulties inherent in hand-description of intonational contours. FDA refers to a set of analysis techniques that extend classic statistical tools to the domain of functions (Ramsay & Silverman [11]). This means that the input elements of the analysis are not (vectors of) numbers but entire curves, represented in terms of functions. In this work we will use one of these tools, namely the functional extension of Principal Component Analysis (PCA). Functional PCA allows the extraction of a compact description of the main shape variations (or Principal Components, PCs) that are present within a dataset of curves, F0 contours in this case. For example, a PC could show that a hump is present in all curves of a dataset, but that this hump varies in amplitude and/or position across the dataset. The original curves are associated to the PCs by means of so-called PC scores., which quantify where in the continuum of the shape variation described by each PC a specific curve is located. For example, very pronounced humps would get a PC score far from zero associated with the PC describing the hump amplitude. Since PC scores are numbers, they can be easily used in further analyses, including being correlated with manual labels (as already shown by Gubian et al. [12]). However, since they are also directly related to shape variations, they supply a means of filling the gap between shape description and quantitative analysis.

In this study, we apply FDA (functional PCA) to a set of prenuclear pitch accents taken from a production study on Standard Southern British English (SSBE). The study is primarily exploratory, in that we will use FDA to provide a descriptive analysis of a number of prenuclear contours. We will show how the observations made on the basis of FDA can be related to the descriptions provided by two autosegmental accounts mentioned above, and also suggest how the use of FDA might begin to identify some possible new avenues for investigation which have been less studied in the past.

2. Methodology

The recordings came from a data set collected by the first author as part of a larger study of the production of intonation in SSBE (Zellers [13]). Eleven female and five male native speakers of SSBE were recorded reading aloud a narrative text that had been written to control for specific segmental and discourse context effects. From this body of recordings, tokens of the speakers producing one target word, 'Emory' (a character's name) were extracted. All of the instances used in this study were prenuclear accents; they were the first pitch accent in an intonational phrase, always followed by at least one other pitch accent, and as such none of them bore focus. The target word was always either the first word in the sentence or else it was preceded by a two-syllable anacrusis. The tokens had been labeled as rising (L*H) or falling (H*L) according to the annotation scheme proposed by Gussenhoven ([4], [5]) for British English. We expect that these categories will roughly coincide with the Pierrehumbert/ToBI L*H and

(L+H)* (cf. Gussenhoven & Rietveld [14]), although the overlap may not be exact; the extent to which these categories map is a question that is beyond the scope of the current work. The distribution of H*L and L*H in the anacrusis or initial categories did not vary significantly (cf. Zellers et al. [15]), although we expect that some within-category variation will arise from this difference in segmental structure (Nolan & Farrar [16]).

In total, 126 tokens of the target word were used in the study. Of these, 56 were produced with three syllables, /'ɛməɹi/ or /'ɛmɜɹi/, and the other 70 were produced with two syllables, /'ɛmɹi/. Using Praat (Boersma & Weenink [17]), the pitch contours were extracted and given an initial smoothing by visual inspection by the first author to remove any octave errors or other spurious pitch points or jumps. The original set of data points comprised 182 tokens; however, 56 items were excluded on the basis that the stressed vowel was produced with initial creak, and therefore the F0 data was unavailable or spurious for part or all of the vowel. All of the items used in the final analysis were produced with modal voicing (i.e. modal for the speaker) over at least the last 67% of the stressed vowel. This cut-off was chosen on the basis of evidence that the vowel onset may already be a location where perceptual saturation occurs (cf. House [18]), and that listeners may therefore disregard vowel-onset F0 information. In this way we are able to retain tokens which still have potentially perceptually useful F0 information while accounting for the fact that the whole vowel is not necessarily produced with modal voice and that therefore the pitch information is not always available to a listener.

Before applying any FDA tool, a standard data preparation procedure has to be followed (for details see Ramsay & Silverman [11] or the website maintained by the second author [1]). First, all sampled F0 contours must be converted to semitones and their average over time subtracted. This helps the automatic extraction of shape-related features by removing global variations, which in our case are likely to be associated with speaker identity (e.g. gender). Then, all sampled curves must be interpolated using the same function basis, often a B-splines basis, as in the current study. From this step onwards all statistical tools are applied to the functional representations.

An unavoidable side effect of the use of a common function basis is that all curves must be defined on a common time interval. However, signals like F0 contours extracted from different realization of the same word obviously have different durations. As a consequence, the functional representations of the original contours must be distorted (or registered) in time in such a way that they exhibit the same duration. To make this process less detrimental, points that have the same meaning across the curve set can be used as landmarks and get automatically aligned in time across the registered functions set. We used the beginning of the last syllable /ri/ as common registration landmark.[2]

After data preparation, functional PCA was applied to the 126 registered functions. All FDA operations were carried out using the freely available R package 'fda' (Ramsay et al. [19]).

[1] Available http://lands.let.ru.nl/FDA

[2] From an intonology point of view, using the end of the first (stressed) syllable as an alignment point would have been preferable. However, given the mixed syllable structure of the tokens in the study, the boundary before the final syllable /ri/ was chosen as a compromise allowing all tokens to be represented together while still maintaining some segmental alignment information.

3. Results

The results of functional PCA applied to our F0 contours show two Principal Components (PC1 and PC2) that contribute significantly to the shape of the F0 contours. Of these principal components, PC1 explains 83.5% of the variance in the contours, while PC2 contributes an additional 9% of explanatory power. Figure 1 shows PC1 and PC2 as variations with respect to the mean. The x-axis represents normalized time and the y-axis pitch in normalized semitones. The mean function is the function obtained by averaging all curves in the dataset at each time point, and it is represented with a solid curve in both panels. PCs are represented with pairs of curves, the '+' and the '-', which show the typical shape of a curve whose PC score is positive or negative, respectively. In fact a PC is a 'correction function', i.e. a function added to or subtracted from the mean function in a proportion given by the PC score associated to a specific original contour. The specific PC score used in the '+' and '-' curves is chosen in a way to facilitate visualization of the shape variation continuum as it corresponds to the variation of PC scores.

Therefore, in our case a positive value for the PC1 score gives a pitch contour that is low at the beginning, and increases across time to be high at the end of the word. On the other hand, a negative value for the PC1 score gives a pitch contour that peaks very early, near the end of the stressed syllable, and then drops throughout the rest of the time-course to end (relatively) very low (recall that the curves show relative values, since time averages have been subtracted).

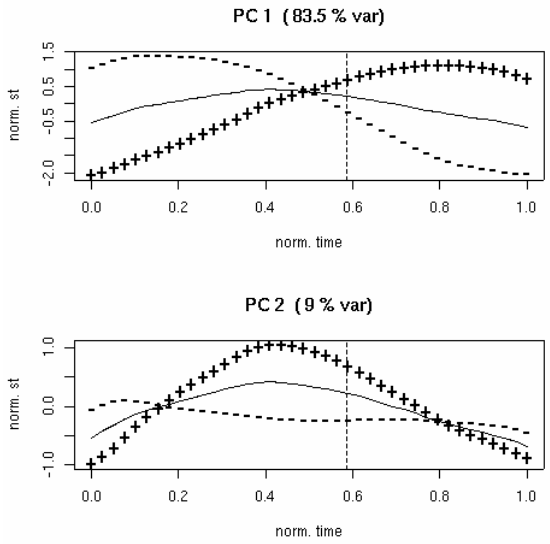

Figure 1: PC1 and PC2 in relation to the mean contour. The vertical bar marks the beginning of the syllable /ri/.

PC2 is more difficult to interpret in the current data. A case-by-case analysis revealed that a few curves labeled as rising that present a relatively late maximum are associated with a PC1 score close to zero (a relatively flat curve) corrected by a high positive PC2 score, which produces a maximum roughly in the middle of the (normalized) time window. Moreover, very rapidly falling contours are associated with negative PC1 and PC2 values, the latter helping to approximate a steep falling contour. Thus in our case PC2 is to be considered a correction function that does not help into discriminating global rise/fall trends but operates mostly into improving the approximation of a small part of the whole dataset, which PC1 does not account for properly. Therefore, in our further discussion, we will focus only on the variation in PC1.

In addition to these qualitative observations, functional PCA allows us correlate PC scores to the original manual labels. Figure 2 shows how the PC1 scores are basically divided into two separate sets by the rising/falling labels at around the zero point. This suggests that what was manually labeled matches with an intrinsic dynamic variation in the curve dataset.

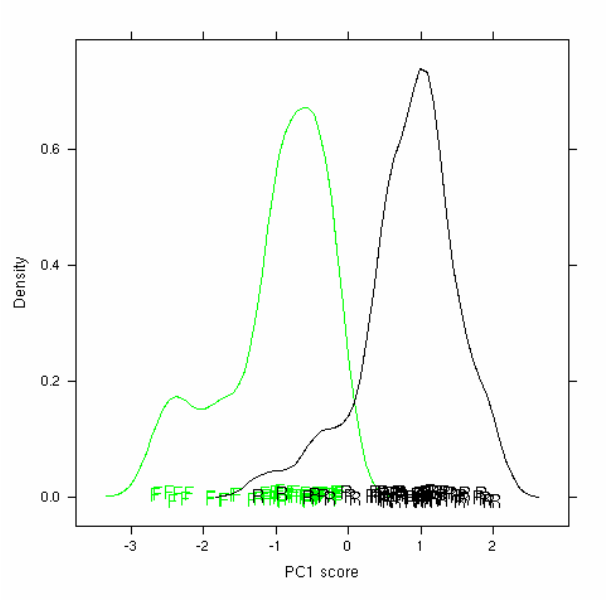

Figure 2: PC1 score densities for contours labeled as rises (R - black) and falls (F - green). The labels were determined by hand-labeling by the first author, independent of the distribution of the contours given by the Functional PCA.

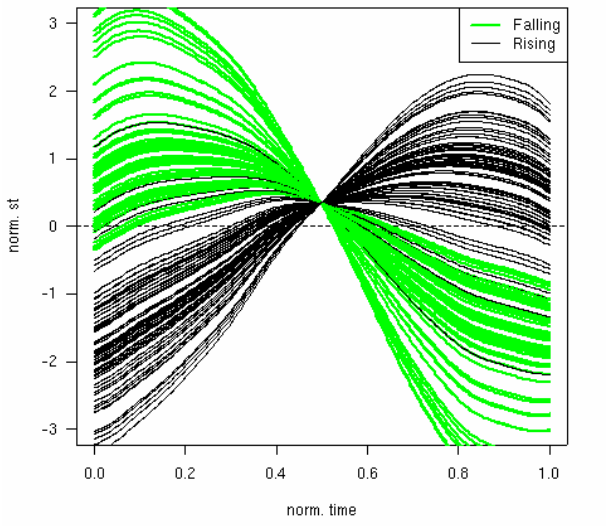

Figure 3: Contours reconstructed using PC1 values

Figure 3 helps to link the PC1 scores back to their meaning in terms of shapes. It shows the approximated representation of the whole dataset that can be obtained by considering only the information provided by PC1. Rising/falling labels are represented in different colors. It is evident that what has been manually labeled as rising/falling are curves that according to the first PC, which explains most of the variance present in the dataset, do indeed have a global rising or falling shape, although there are a few exceptions, which may be due to contextual variations that are not apparent in the PC1-based reconstruction. It is in these cases in particular that taking PC2 into account becomes more helpful for the accurate reconstruction of the contours. Recall that the labels 'rising' and 'falling' were not used by the FDA analysis at any point; in other words, the relative similarity among different contours was determined without reference to the manual labeling.

4. Discussion

The '+' and '-' curves in Fig 1 give a starting point for investigating the 'important' features of the F0 variation. In the positive case, F0 starts low and rises across the course of the word, ending high (apparently with an F0 peak in the third syllable). In the negative case, F0 starts fairly high and rises slightly to a peak at the end of the first syllable, then drops quickly throughout the course of the word to end low. If we plot the distributions of PC1 scores associated with contours which were labeled as H*L or L*H in the initial analysis as in Figure 2, we see a clear distributional difference, although it is important to note that this in and of itself does not guarantee a phonological boundary, but simply a difference in distributions. In this case, by using FDA, we may begin to make a more principled phonetic distinction between two accent categories which have already been identified in the intonational phonology. The criterion for this distinction is the function's PC1 value, which we could begin to interpret in a very general way as the global direction of the contour: rising or falling.

However, by using FDA it is possible to observe more features of the contours than simply the global movement direction. For example, in Fig. 3 in the contours with a negative PC1 value, the pitch peak occurs early, likely near the boundary of the stressed (first) syllable, and is preceded by a small rising movement, though the size of the rise to the pitch peak is not at all comparable to the size of the fall, which ends at a similar level to the low onset of the positive-PC1 contours. Another observation that we may make is that the peak of the positive-PC1 contours appear to be lower than the peak of the negative-PC1 contours. Recalling that the mean over time was removed from the original contours (Sec. 2); this means that negative-PC1 contours tend to exhibit ampler excursions than positive ones.

Using the mean contours produced by FDA, we are able to observe an asymmetry that is not necessarily apparent in either the simple descriptions 'rising' and 'falling' as in the British tradition, or in the description of contours as interpolating between high and low tones, as in the autosegmental approach; not only the global difference between high and low pitch, but also a difference in the way the transition is made. This could have further consequences for a phonetically-informed phonological model; for example, the ToBI model (Silverman et al. [1]) describes both of the contour categories presented above as rises (L+H* and L*+H) on the basis of the rising pitch movement on the stressed syllable, with the contrast being a difference in alignment of the high peak with the segmental stream. However, using FDA, it is possible to see that the pitch rises vary in more than the alignment of the peak; the pitch excursions differ greatly[3], as do the shapes of the pitch movements themselves. The extent to which this phonetic detail reflects differences in the

[3] Note that almost none of the rise slopes in this study approached the values calculated by Xu & Sun ([20]) as articulatory limits for SSBE speakers; this means that the size of the pitch excursion is unlikely to have been determined by articulatory factors alone, even though the time available for the pitch movement was more limited in these cases.

phonological interpretation of the contours is yet to be determined, although current research in German and Italian (Niebuhr [21], D'Imperio et al. [22]) has shown that changing the shape of a contour from a peak to a plateau can have strong perceptual consequences in terms of the identification of the pitch accent. Of course, the autosegmental approach, despite using H and L as symbols to represent intonational contrasts, does not limit the investigation of intonation to only pitch turning points. FDA is a way of beginning to come to grips with other phonetic dimensions of the speech signal (F0 and otherwise), regardless of how we ultimately model this variation.

One of the benefits of the analysis provided by FDA is that it lends itself to pitch synthesis as well. That is, it is possible to use the results of the FDA analysis to synthesize different pitch contours on the basis of the PCs, which can then be used in perceptual testing (cf. Gubian et al. [10]). Again, the use of FDA can overcome the difficulties posed by hand-synthesizing stimuli by identifying the important characteristics of whole contours, and transferring those to experimental stimuli. This makes analysis by FDA even more valuable to intonologists in the sense that its models can be directly applied to and validated by perceptual testing.

5. Conclusion

We have applied the methods of Functional Data Analysis to a set of pitch contours to show how FDA can be useful to researchers in intonation by providing a less theory-dependent characterization of which aspects of a contour are most stable across many tokens. Although FDA cannot on its own make any statements about what is meaningful in the F0 contour, it enables the possibility to identify further similarities across many contours which may have thus far gone unnoticed, or might at least have been uninterpretable up until now. This makes FDA a powerful tool for intonation researchers as they seek to characterize meaningful variation in F0 and describe the systems underlying that variation.

6. Acknowledgments

Many thanks to Lou Boves and Oliver Niebuhr for helpful discussion on this topic. This research was supported by the EC Marie Curie Training Network *Sound to Sense* (MRTN-CT-2006-035561), and by the ESRC First Grant *Categories and gradience in intonation* (RES-061-25-0347).

7. References

[1] Silverman, K., Beckman, M., Pitrelli, J., Ostendorf, M., Wightman, C., Price, P., Pierrehumbert, J. & Hirschberg, J. ToBI: A standard for labelling English prosody. In *Proceedings of the 1992 International Conference on Spoken Language Processing*, pp. 867-870, 1992.

[2] Pierrehumbert, J. *The Phonology and Phonetics of English Intonation*. Doctoral dissertation, MIT, 1980.

[3] Beckman, M. and J. Pierrehumbert. Intonational Structure in Japanese and English. *Phonology Yearbook III*, pp. 15-70, 1986.

[4] Gussenhoven, C. *On the Grammar and Semantics of Sentence Accents*. Dordrecht: Foris, 1984.

[5] Gussenhoven, C. *The Phonology of Tone and Intonation*. Cambridge: Cambridge University Press, 2004.

[6] Cruttenden, A. *Intonation*. Cambridge: Cambridge University Press, 1986.

[7] 't Hart, J., Collier, R., & Cohen, A. *A perceptual study of intonation. An experimental- phonetic approach to speech melody*. Cambridge: Cambridge University Press, 1990.

[8] Knight, R.A. *Peaks and Plateaux: the production and perception of high intonational targets in English.* Doctoral dissertation, University of Cambridge, 2003.

[9] Syrdal, A.K. & McGory, J. Inter-transcriber reliability of ToBI prosodic labeling. In *Proceedings of International Conference on Spoken Language Processing*, Beijing, China., pp. 235-238, 2000.

[10] Gubian, M., Cangemi, F. & Boves, L. Automatic and data driven pitch contour manipulation with Functional Data Analysis. To appear in *Proceedings of Fifth International Conference on Speech Prosody*, Chicago, USA, 2010.

[11] Ramsay, J.O. & Silverman, B.W. *Functional Data Analysis* (2nd Ed.) New York: Springer, 2005.

[12] Gubian, M., Torreira, F., Strik, H. & Boves, L. Functional Data Analysis as a tool for analyzing speech dynamics: a case study on the French word *c'était*. In *Proceedings of 10th Interspeech*, Brighton, UK, pp. 2199-2202, 2009.

[13] Zellers, M. *Prosodic detail and topic structure in discourse*. Doctoral dissertation, University of Cambridge. In preparation.

[14] Gussenhoven, C. & Rietveld, A.C.M. An experimental evaluation of two nuclear tone taxonomies. *Linguistics* 29, pp. 423-449, 1991.

[15] Zellers, M., Post, B. & D'Imperio, M. Modelling the intonation of topic structure: two approaches. In *Proceedings of 10th Interspeech*, Brighton, UK, pp. 2463-2466, 2009.

[16] Nolan, F. & Farrar, K. Timing of F0 peaks and peak lag. In *Proceedings of ICPhS San Francisco*, pp. 961-964, 1999.

[17] Boersma, P. & Weenink, D. *Praat: doing phonetics by computer* [Computer program]. Version 5.1.31, retrieved 4 April 2010 from http://www.praat.org/

[18] House, D. *Tonal Perception in Speech*. Lund: Lund University Press, 1990.

[19] Ramsay, J.O., Hookers, G. & Graves, S. *Functional Data Analysis with R and MATLAB*. New York: Springer, 2009.

[20] Xu, Y., & Sun, X. Maximum speed of pitch change and how it may relate to speech, *JASA 111*, pp. 1399-1413, 2002.

[21] Niebuhr, O. Alignment and pitch-accent identification – implications from F0 peak and plateau contours. Submitted to *Speech Communication*.

[22] D'Imperio, M., Gili Fivela, B., & Niebuhr, O. Alignment perception of high intonational plateaux in Italian and German. To appear in *Proceedings of Fifth International Conference on Speech Prosody*, Chicago, USA, 2010.